BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 9: e181523
doi: 10.3897/biss.9.181523

OPEN ACCESS

Conference Abstract

The Impact of Crowdsourcing Projects in the Digitisation Effort of European Herbaria

Giulia Micai[‡], Wouter Addink[§,|], Anton Güntsch[¶], Heimo Rainer[#]

‡ Natural History Museum of Vienna, Vienna, Austria
§ Naturalis Biodiversity Center, Leiden, Netherlands
| Distributed System of Scientific Collections - DiSSCo, Leiden, Netherlands
¶ Freie Universität Berlin, Berlin, Germany
# Natural History Museum Vienna, Vienna, Austria

Corresponding author: Giulia Micai (giulia.micai@nhm.at)

Received: 05 Dec 2025 | Published: 09 Dec 2025

Citation: Micai G, Addink W, Güntsch A, Rainer H (2025) The Impact of Crowdsourcing Projects in the Digitisation Effort of European Herbaria. Biodiversity Information Science and Standards 9: e181523. https://doi.org/10.3897/biss.9.181523

Abstract

The digitisation of natural history collections is nowadays among the most critical challenges and strategic efforts faced by museums and scientific institutions around the world.

In the frame of the EU-funded project TETTRIs (Transforming European Taxonomy through Training, Research and Innovations), we built upon existing platforms for specimen annotation, such as Les Herbonautes (Muséum national d'Histoire naturelle, Paris), Die Herbonauten (Botanischer Garten und Botanisches Museum Berlin) and DoeDat (Plantentuin Meise), to extract new data from botanical specimens physically deposited in various European herbaria and increase the engagement of the general public with collections.

The objectives of this work have been to expand the use of these pre-existing tools to a broader community of citizen scientists; to reach different linguistic communities; and to get the platforms to be used in more countries. We also investigated the connection of these tools with the new DiSSCo Digital Specimen infrastructure (DiSSCOver). DiSSCOver enables annotations to be made directly on the specimen data connected through this infrastructure.

© Micai G et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

The digitisation of herbarium specimens significantly enhances access to the collections for researchers and institutions located all over the world and facilitates large-scale studies. Furthermore, digitisation helps to preserve the physical specimens by reducing the need for handling and transportation. It also allows novel uses of metadata in combination with digital material: for example, herbarium label data have been used to study shifts in plant phenology and also to generate species distribution models (since accurate models rely on large numbers of data points that can efficiently be achieved via the repositories of digitized herbarium records).

Through the joint use of the three aforementioned crowdsourcing annotation platforms, we have been able to reach various communities of volunteers who helped overcome linguistic barriers that might make data extraction difficult on a local scale. They have made the information on the labels machine-readable, searchable and safely stored. Relying on a broad, varied and knowledgeable volunteer community allowed us to overcome the difficulties that are typical of label transcriptions. One of the main challenges has been the transcription of difficult handwriting or old, obsolete scripts. Also, we successfully overcame linguistic barriers by delivering specimens otherwise hard to annotate on a local scale to the corresponding linguistic community. Thanks to the knowledge of amateur specialists it has been possible to correctly attribute locations in presence of, for example, historical names of locations and to fill up incomplete data. By involving a large number of volunteers and processing high numbers of specimens in a limited time, the process of transcription sped up drastically.

Since the beginning of the project, we have been able to launch and complete nine digitisation campaigns: three on Les Herbonautes, three on Die Herbonauten and three on DoeDat. By doing so, we have been able to transcribe the label data of 5,603 specimens written in 3 different languages, involving approximately 115 volunteers.

In particular, a special effort was put into delivering specimens with labels written in German, French and Dutch languages to the corresponding linguistic communities: for example, labels in French but specimens conserved at Berlin have been transcribed online by a French-speaking community of volunteers, and so on, overcoming the linguistic barrier that kept this material from being transcribed. This action contributed to creating an extended community of volunteers involved in more than one annotation platform at the same time, promoting a knowledge exchange across national borders.

Keywords

specimen annotation, citizen science, annotation campaign, label transcription

Presenting author

Giulia Micai

Presented at

Living Data 2025

Acknowledgements

Thanks to the platform hosts of the crowdsourcing annotation initiatives DoeDat, Les Herbonautes, and Die Herbonauten for granting access to their data and for their support in facilitating this research.

Funding program

This work was supported by the TETTRIs project, which receives funding from the European Union's Horizon Europe Innovation Actions under grant agreement No. 101081903.

Hosting institution

Natural History Museum of Vienna

Conflicts of interest

The authors have declared that no competing interests exist.